



## The co-consciousness hypothesis

Frédérique de Vignemont

### ► To cite this version:

Frédérique de Vignemont. The co-consciousness hypothesis. *Phenomenology and Cognitive Science*, 2004, 3, pp.97-114. ijn\_00169608

**HAL Id: ijn\_00169608**

**[https://hal.science/ijn\\_00169608](https://hal.science/ijn_00169608)**

Submitted on 4 Sep 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## The co-consciousness hypothesis

FRÉDÉRIQUE DE VIGNEMONT

*Institute of Cognitive Neuroscience, University College London, UK*  
(E-mail: f.vignemont@ucl.ac.uk)

**Abstract.** Self-knowledge seems to be radically different from the knowledge of other people. However, rather than focusing on the gap between self and others, we should emphasize their commonality. Indeed, different “mirror matching mechanisms” have been found in monkeys as well as in humans showing that one uses the same representations for oneself and for the others. But do these shared representations allow one to report the mental states of others as if they were one’s own? I intend in this essay to address the epistemic problem of other minds by developing Ayer’s notion of co-consciousness.

**Key words:** Ayer, co-consciousness, empathy, mirror neurons, theory of mind

### Introduction

In the *Meditations*, Descartes (1641) raised the problem of other minds by wondering whether the people he was seeing in the street were only bodies or whether they had also minds. This problem may be articulated in two different ways. The skeptical version challenges the very possibility of knowing the minds of others and arises from the Cartesian conception of mentality: a mental state can be understood only from a first-person perspective. The epistemic version investigates the nature of the knowledge of other minds and its relation with self-knowledge. While I have a privileged direct knowledge of my own mental states, I do not have any immediate access to other minds, but only to public behaviors and I can only infer what the other people think from what I can observe. Thus, it seems that the access I have for my own mental states is radically different from the access I have to the states of others. However, as Davidson (1987) points out, if the same mental concept is applied sometimes on the basis of observational evidence, and sometimes on the basis of none, then the concept is ambiguous. In other words, what is common between my pain, which I immediately know, and your pain, which I infer from the fact that you are crying? On what basis could I compare them? In order to avoid such problems, some philosophers have tried to reduce the gap between self-knowledge and the knowledge of other minds. Ryle (1966) for instance denies the existence of a privileged access to one’s own mental states and suggests that self-knowledge is based on observational and inferential support in much the same way as knowledge of others: “No metaphysical Iron Curtain exists

compelling us to be forever absolute strangers to one another” (Ryle 1966, p. 31).

Similarly, I will argue for a common epistemic basis of self-knowledge and the knowledge of other minds, but unlike Ryle I will not reduce the former to the latter.

### **The co-consciousness hypothesis**

I do not intend here to address the whole issue of the problem of other minds. Rather, I would like to question the radical difference usually assumed between self-knowledge and knowledge of other minds by appealing to the notion of “co-consciousness” as described by Ayer (1963). More particularly, I will answer to the following question: Can I be aware of mental states of others as if they were my own?

#### *Theory of mind from a first-person and a third-person perspectives*

Nowadays, philosophy of mind and developmental psychology investigate the underlying mechanisms of the so-called “theory of mind”. By theory of mind, I mean the general ability to recognize mental states and to predict behaviors. More precisely, theory of mind can be from a first-person perspective (how do I know my mental states?) and from a third-person perspective (how do I know someone else’s mental states?). I will briefly characterize two main views of theory of mind, which more or less emphasize the difference between the two perspectives.<sup>1</sup>

The Simulation Theory (ST) holds that we represent the mental activities of others by mentally simulating them, that is, by generating similar activities and processes in ourselves (Goldman 1993a; Gordon 1996). In other words, I know the other by putting myself into her shoes. ST thus focuses on the theory of mind from a third-person perspective and does not provide any particular reply to the problem of the first-person perspective. We may yet mention Goldman (1993a, 1993b), who claims that I know the other by simulation and that I know myself by a kind of inner sense. Consequently, there is a primacy of the self (which is directly known) over the other (known by simulation on the basis of one’s own states). The challenger to ST is the so-called “theory” theory (TT), which assumes that one has a body of general knowledge about the mind, rather like a scientific theory (Gopnik 1993). I know what I think as well as what someone else thinks by appealing to a tacit functional theory of the mind. There would be only a difference of degree between the knowledge of the self and the other, because I apply the same theory in both cases.

Some experimental results tend to argue in this latter direction. Nisbett and Wilson (1977) show that we are not aware of our problem-solving processes

and that we compensate this lack of insight by rationalizing *a posteriori* the reasons of our behavior using cultural or personal implicit theories. Gopnik (1993) also demonstrates the symmetry of the development of theory of mind for the first-person and for the third-person perspectives. Children acquire at the same time knowledge of their own mental states and knowledge of the mental states of others and they show the same difficulties in both cases. For instance, 3-year-old children have difficulty understanding the fact that beliefs could be false. In one experiment, when they open a closed candy box, it turns out that there are pencils inside it, rather than the candy they have been expecting. They predict after that that the other person will think that there are pencils there even if she did not see the content of the box (Wimmer and Perner 1983). They make the same error when they are asked about their own immediately past false beliefs: they report that they already thought that there were pencils in the box (Gopnik and Astington 1988). Therefore, according to Gopnik, we do not have any privileged access to our own states that we do not have for the states of others. Like the chess master who feels as if he directly perceived the strength of his game, we feel as if we had an immediate knowledge of ourselves, but we are the victims of a phenomenological illusion, which results from our level of expertise in the grip of the tacit theory of the mind.

One may comment on the experimental validity of Gopnik's results. However, I will rather focus on her philosophical view and distinguish three independent claims. The first one is that self-knowledge is not infallible, as it is confirmed by phenomena such as self-deception or blindsight, and nobody will deny that except perhaps Descartes. The second assumption is more controversial: in every case, for determining my mental states, I apply functional laws that I master perfectly well. However, even if I sometimes theorize and confabulate about myself, it does not seem plausible that I know that I am in pain because I infer my pain from my behavior. The third claim is the most relevant one for us: self-knowledge and the knowledge of other minds may have something in common, or in other words, self and others are not so far from each other at the epistemic level. I will develop this latter hypothesis independently from the other claims. I will argue that we do not need to assume a radical distinction between knowledge from a first-person and from a third-person perspectives. Yet, it does not imply that we have to postulate the existence of a shared tacit scientific theory of the mind. There may be another possibility, which I will describe by using Ayer's notion of co-consciousness.

### *Co-consciousness*

Privacy is often assumed as the mark of the mental. Ayer (1963) explains this idea as follow:

Many philosophers would conclude that this was so, for the following reason: they would argue that being directly aware of someone's, say Mr. X's, thoughts or feelings entailed having them, and having Mr. X's thoughts or feelings entailed being Mr. X, and that it was logically impossible that anyone should be Mr. X, but Mr. X (Ayer 1963, p. 65).

One possible interpretation of the privacy hypothesis is that we cannot share our thoughts and feelings: I can be conscious only of my own states from a first-person perspective. However, Ayer argues against the privacy hypothesis by showing the logical possibility of what he calls "co-consciousness<sup>2</sup>." One person is said to be co-conscious by

being able to make reports about the mental states of others in the same 'immediate' way as one makes reports about one's own, to report on them, in short, as if they were one's own (Ayer 1963, p. 65).

In other words, a person is co-conscious with someone else if she is conscious of the same experience with the other person. We may notice an ambiguity in this definition. Indeed, we can provide two different interpretations of the idea of "same experience" depending on whether we consider numerical or qualitative identity.

#### *Narrow interpretation*

*a* is numerically identical to *b* if *a* and *b* are one and the same entity. Therefore, one person is co-conscious with someone else if she is conscious with the other person of the very same mental token. Co-consciousness thus requires having access from a first-person point of view to someone else's states, that is, to be directly aware of them independently of any inference from her behavior. If so, I am able to report the mental states of others as if they were my own because I have the same kind of knowledge as for my own states. The best example of co-consciousness in the narrow sense would then be the phenomenon of telepathy: one directly "reads" in the mind of another person, in the same way as one reads in one's own mind. Similarly, we could imagine being connected by a neural cable to someone else's brain. For instance, Ramachandran and Hirstein (1997) suppose that Mary, Jackson's scientist who has never seen any color,<sup>3</sup> would be able to know what it is like to see colors by being properly (directly and neurologically) connected to the brain of someone who sees a red patch and by sharing his red visual experience. We could also imagine connecting the cable not to the visual cortical area, but to the somatosensory area in such way that Mary could internally feel the bodily sensations of the other.<sup>4</sup>

In all these cases, self-knowledge is no longer privileged by comparison to the knowledge of others. Then how could they be distinguished? In other words, how can the subject determine which are her own states and which are the states of others? Ramachandran and Hirstein conclude that Mary would know the quale of red. Thus, she would have the experience of red and would

report this experience as if it was her own despite the fact that she does not actually see the red patch. In traditional telepathy stories, there is no confusion between people's thoughts. However, if we have strictly the same access to the thoughts of others and to our own thoughts, then we would have no epistemic cue for attributing them to their actual thinker. Co-consciousness interpreted in a very narrow sense might lead to a mental chaos without any frontier between oneself and the others. This problem could be avoided if we interpret the co-consciousness hypothesis in a wider sense by appealing to the notion of qualitative identity.

*Wide interpretation*

*a* is qualitatively identical to *b* if *a* and *b* share the same properties, but are instantiated by two different items. Therefore, one person is co-conscious with someone else if each of the two persons is conscious of a token mental state that is qualitatively identical to a mental state of which the other is conscious.<sup>5</sup> If so, I am able to report the mental states of others as if they were my own because I have an immediate access to my own states, which are properly similar to their mental states. However, not just any sharing phenomenon can be considered as co-consciousness and we need to specify in more details its necessary conditions. It does not suffice to share the same state because it can be a mere coincidence. For instance, we would not interpret the fact that two persons entertain similar thoughts without any relation between them as an illustration of co-consciousness.

Suppose that both Peter and Jack see that Peter's legs are crossed. They have the same kind of access to Peter's bodily state, but there is no direct causal link between Peter's and Jack's visual state. Jack is not necessarily able to report Peter's visual state as he does not necessarily know what Peter is actually looking at and thus does not know what is in his mind.<sup>6</sup> Therefore, Jack is not co-conscious with Peter. Co-consciousness has first to be understood as a kind of knowledge of the mind of the other. It implies not only the sharing of a certain mental state, but also the knowledge of this sharing: Jack is co-conscious with Peter if he knows that he has the same perceptual experience as Peter, and thus knows Peter's perceptual experience. Furthermore, co-consciousness requires more than mere simultaneity, it requires causality: Jack has the perceptual experience because Peter has it. The phenomenon of joint attention (Baron-Cohen 1997) could fit this description because Jack would then look at Peter's legs because he would have oriented his attention toward where Peter is actually looking. However, the two visual states would not be strictly similar, in either their content or in their functional role. Peter and Jack do not have the same spatial perspective and only Peter's visual state of his legs is functionally related to the actions that he could perform. A strict qualitative identity would involve a similarity at the intentional level (the same attitude with the same content), at the functional level (the same

causal role) and at the phenomenological level (the same associated qualia). However, co-consciousness does not imply a perfect “copy” of mental states of other people, but only the sameness of the state on the relevant aspects, which allows one to report the state correctly. For instance, even if my experience of sadness for your father’s death is less intense than yours and will not lead me to cry the whole day, I can still know how you feel by this partial sharing.

To sum up, co-consciousness in the wide sense requires (1) two distinct persons;<sup>7</sup> (2) a causal relation between the two states such that one has the state because of the other having it; (3) a certain isomorphism (intentional, functional and phenomenological) of the two mental tokens that leads to an accurate knowledge of the states of the other.

In order to challenge the privacy hypothesis, we may maintain either like Gopnik that self-knowledge is nothing more than a specific type of knowledge of the others or conversely like Ayer that there is a way of knowing the other that is similar to the way of knowing oneself. I have tried here to develop this latter hypothesis by describing two interpretations of co-consciousness. I have shown that the narrow interpretation may lead to a total lack of discrimination between self and other. The wide interpretation seems more promising, but also raises some problems. We may indeed wonder if it does not leave the epistemic gap too wide. The three conditions described do not suffice for accounting for the knowledge of others: I may indeed be in a certain state that is caused by someone else’s state, which is qualitatively similar, without yet knowing that the other has the same state. Consequently, I also need to be able to recognize that the other has a certain state from the fact that I have it. However, as we will see, this last condition is problematic.

### **From empathy to mirror neurons**

Till now, I have developed one possible reply to the epistemic problem of other minds by specifying the necessary conditions of co-consciousness. Ayer discusses the question of the privacy of the mental at the level of logical possibility, without entering into the problem of what is empirically possible. However, it would be interesting to see whether there is any actual phenomenon that could argue for the empirical plausibility of co-consciousness. Furthermore, it would shed a new light on the relation between self-knowledge and the knowledge of other minds. I will start with the most salient example of co-consciousness, that is, empathy.

We have all at some time cried with the hero of a movie and laughed just because our neighbors were laughing. The empathic knowledge of other people is ambiguous, as it is not strictly identical to self-knowledge nor to the usual inferential knowledge of the others. Indeed, in empathy I do “feel

with” the other her emotion, and in this sense it can be interpreted as a case of co-consciousness.

The literature tends to confuse empathy and sympathy. However, it is important to clearly distinguish them. For instance, I learn that my friend’s husband is betraying her. I feel sad for her even if she does not know her husband’s betrayal. Thus, my emotion does not match with her emotion. I do sympathize for her. By contrast, I see my friend crying because her husband is betraying her. I may cry with her, feeling as if I was in her situation. I do empathize with her. In brief, sympathy and empathy play different roles: the former constitutes a social link while the latter constitutes a way of knowing the others (Wispé 1986). I will thus focus on empathy, and more precisely on emotional empathy. Empathy is indeed a continuum that includes the following:

- Emotional contagion, which relies on a direct link between the perception of the emotions felt by the other and one’s own experience of the emotion. For instance, the vision of my friend’s smile makes me automatically smile and triggers a feeling of happiness.
- Emotional empathy, which involves the feeling of someone else’s emotion and the understanding that the other feels the same emotion, like in the previous example: I feel sad with my friend because of her husband.
- Cognitive empathy, which also involves the knowledge of the emotion of other people, but which stays more detached from the emotion. Wispé (1986) reports for instance that during the Second World War, Nazi officers used to put themselves in the shoes of their victims in order to imagine what would be the most frightening for them, but they were not scared for all that.

Emotional empathy may be interpreted as an illustration of co-consciousness, but not in the same sense as telepathy. Unlike telepathy, empathy does not indeed rely on the direct “reading” of the emotion felt by the other. It involves the process of theory of mind. We are then confronted to two possibilities. Suppose that you learn that your friend has just missed his flight. How do you know what he feels? Do you use a functional law such as “when you do not get what you want, you’re really frustrated and angry”? Or do you put yourselves in his shoes and imagine what you would feel if you had missed your flight?

*“Theory theory” interpretation of empathy*

According to the theory theory (TT), the subject determines the target’s mental states by appealing to a tacit theory of how the mind works, which is partly based on the subject’s past experiences and on the laws he has drawn from them:



It is likely that people often empathize not because they have put themselves cognitively in another's place, but because they have retrieved relevant information from their memories (Eisenberg and Strayer 1987, p. 9).

However, what kind of memory do we appeal to? As TT consists in an abstract theoretical knowledge of functional laws, we should assume that empathy should be based on semantic memory, which stores one's own past experiences from a third-person point of view. Contrary to episodic memories, semantic ones are detached from any emotional connotation and could not produce the feeling of anger associated to your understanding of what your friend is experiencing (Johnson et al. 1993). Ravenscroft (1998) points out this general problem: if I use a detached theory of other people, I would never be able to feel what they feel and I would be in the same situation as Mary, who knows every objective fact about colors, but not what it is like to see colors. Furthermore, empathy would then constitute nothing more than an inferential indirect knowledge of other minds and could not be considered as a co-consciousness phenomenon.

#### *Simulation theory of empathy*

On the contrary, according to the simulation theory, the subject does not need to master any functional law; all he needs is to simulate his target's perspective on the world and to let his emotional processing system work as usual. Goldman (1992, 1993a) claims that empathy is only a particular case of simulation:

Paradigm cases of empathy... consist first of taking the perspective of another person, that is, imaginatively assuming one or more of the other person's mental states. ... The initial 'pretend' states are then operated upon (automatically) by psychological processes, which generate further states that (in favorable cases) are similar to, or homologous to, the target person's states. In central cases of empathy, the output states are affective or emotional states (Goldman 1993a, p. 371).

Goldman is actually describing only cognitive empathy, which remains detached from the emotion of the other and can be explained by an "off line" simulation. Such explanation could then be subject to the criticism raised by Ravenscroft against the TT interpretation of empathy (Adams 2001). However, as Harris (1989) notices, empathy cannot be reduced to the standard case of simulation because the empathic simulation is not always "off line." Goldman does not provide any account for emotional empathy, which involves the actual feeling by the subject of the emotion of the other. You are not as angry as your friend who has missed his flight, but you still deeply feel the frustration as if you were the person who has missed it. In other words, you become able, by empathy, to report the emotions of others as if they were your own because you also feel them. Emotional empathy seems then to fulfill the criteria of

co-consciousness: (1) two persons, (2) both aware of the same kind of emotion, and (3) the emotion of the subject being elicited by the situation that caused the simulated target's emotion.

*The 'shared manifold' hypothesis*

We may even generalize this analysis to other phenomena: co-consciousness is not indeed limited to emotions and may happen with other kind of mental states. Gallese (2001, 2003) suggests extending the model of empathy to all the aspects of behavior. We can find in various cognitive domains self/other equivalence implemented in cortical areas common to action and perception, emotional experience and emotional recognition, sensation of pain and observation of pain. For instance, Damasio (1999) reports the case of the patient S who is unable to experience the feeling of fear or to recognize it in someone else, even though she keeps an abstract knowledge of what fear is. Calder et al. (2000) also describe the case of a patient with insula and putamen damage who is impaired in both the experience of disgust and the recognition of it in others. Carr et al. (2003) indeed show that both observation and imitation of facial emotions activate the same cortical areas (the premotor cortex, the insula and the amygdala). Furthermore, the same neurons in the human cingulate cortex respond whether a pinprick is applied to the subject himself or to the examiner (Hutchison 1999). Thus, we may postulate that the same representations are involved in the experience and the observation of emotions and bodily sensations. According to Gallese, this "shared manifold" constitutes the basis of intersubjectivity and can be articulated at three levels: at the phenomenological level with empathy, at the functional level with simulative processes, and at the subpersonal level with mirror neurons. I will focus now on this latter level in order to investigate more deeply the co-consciousness hypothesis.

*The dual interpretation of shared representations*

We have seen that empathy may be understood as a kind of simulation (diagramed in Figure 1).

However, this schema does not account for emotional empathy, that is, for the fact that the subject also feels the emotion E. Furthermore, it omits to

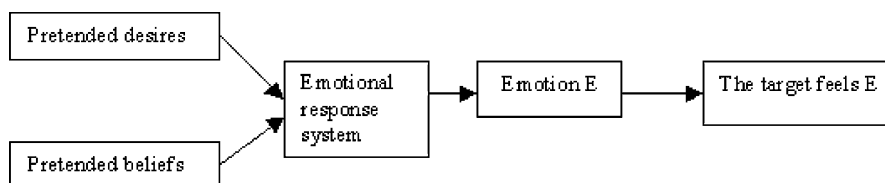


Figure 1. Traditional model of ST.

specify the subject to whom the mental states are ascribed during the simulation process and does not allow any real understanding of co-consciousness. I will thus appeal to the model of action for determining the relation between self-knowledge and the knowledge of others. Since the beginning of the 20th century with Theodor Lipps, empathy has been described as an internal imitation of perceived movements executed by someone else. More recently, Preston and de Waal (2002) suggested interpreting empathy on the “Perception–Action” model: the perception of the emotion automatically activates the representation of this emotion in the subject. This model is based on the existence of shared representations of actions in mirror neurons.

“Mirror matching mechanisms” have been found in monkeys as well as in humans: the same cortical areas – and specifically certain neurons – are activated during observation of other people’s actions and execution of one’s own actions (Rizzolatti et al. 1996; Decety et al. 1997). Thus, doing an action and observing it activate the same internal motor representation. The shared representations of action are “pragmatic” and code actions from the first-person point of view of the agent in interaction with the world (Jeannerod 1994). Such shared representations may underlie the phenomenon of co-consciousness, as Gallese suggests. However, we can give two different interpretations of the role of mirror neurons. On the one hand, mirror neurons fire during observation of actions performed by other, because the subject covertly and implicitly simulates these actions. In this case, mirror neurons represent only the actions performed by oneself. On the other hand, mirror neurons fire during observation and execution of action, because they represent one’s own actions as well as someone else’s actions. In this sense, they are neutral with respect to the agent. I now want to analyze both interpretations in more details.

#### *The primacy of the self*

According to the first interpretation, when someone acts and is observed, the same kind of action representation is activated in the agent and in the observer. Action representations are shared in the sense that different people have the same kinds of representation activated for the same actions. This social sharing could explain how we can understand each other: isomorphism between action representations allows the subject to know the goal of others with the help of his own representations. This is a general phenomenon of “investigation by re-creation” described by Goldman (2002): by observing the other acting, the subject imagines himself performing the same action, a simulation that does not lead to an execution, because it is inhibited. As a kind of covert simulation of the action performed by someone else, mirror neurons would represent only the actions performed by the self. In other words, the agent is always specified and he is always me. Therefore, mirror neurons allow us to “to map the non-self onto the self” (Gallese 2000). There is a primacy of

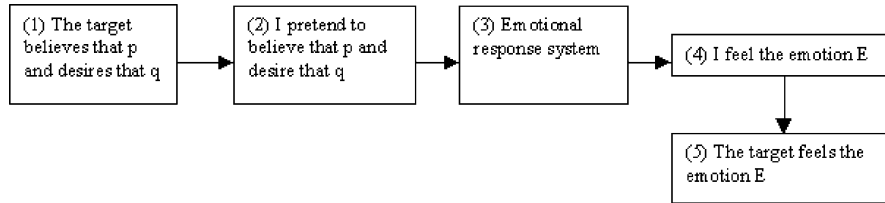


Figure 2. Self model of ST: from me to you.

the self, the other being understood on the basis of the self. We may describe the content of the shared representations in the following hypothesis.

*The Primacy of the Self Hypothesis: Mirror neurons are motor representations with the self as the agent: [I, action, goal].*

We may then provide the same interpretation for empathy as for mirror neurons. According to the hypothesis of the primacy of the self, I understand the others from myself (see Figure 2). Thus, I pretend that I have such desires and such beliefs and that I feel the emotion E and conclude that the target feels E. The lack of inhibition of step (4) explains why I do feel the target's emotion in emotional empathy.

To sum up, the action performed by others is understood from the covert action performed by the self. This description seems to fit perfectly with the definition of co-consciousness in the wide sense. However, we may wonder whether it can reduce the difference between self-knowledge and knowledge of others. As previously said, simulation theory does not pretend to eradicate the epistemic gap between the self and the other. I directly know which state I am in, a state that is shared by the other, but I still need to attribute my state to the other, to infer from my state to the state of the other. Like Gordon (1996), we may wonder about the legitimacy of this inference "from me to you." Empathic knowledge is thus far from being immediate, even if I do not make any inference from the behavior of the other. Though I know the other by knowing myself, two different kinds of knowledge are involved here, one inferential (from the self to the other) and one direct.

Moreover, Gallagher (2001) notices that at the neural level: "perception of action is already an understanding of the action; there is no extra step involved that could count as a simulation routine." I would rather say that there is no extra step for an inference. Gallagher concludes that mirror neurons are the neural bases of direct intersubjective *perception*. Similarly, I want to suggest another interpretation of shared representations that avoids the difficulties involved with making inferences: rather than inferring from me to you, we appeal to neutral representations common to you and me. All we need is to give up the primacy of the self-hypothesis for a more neutral interpretation.

### *A neutral account*

Our starting point was simply that at the neurophysiological level observation and execution of action activate some common areas. Therefore, we can say that the same representation is activated for one's own actions and for actions performed by others. What mirror neurons represent is action. They do not code specific parameters of movements, like the agent's strength or the position of the limbs. For the most part, they represent only general kinds of action. As they do not take into account the peripheral factors of action, they allow establishing a motor equivalence between execution and observation, no matter how different the agent and the observer are. In the same way, they are also partly indifferent to the person who is the agent. Even if they represent actions from a first-person perspective, it does not imply that the agent of the action is oneself. In other words, they do not represent actions performed by the self, but actions performed by one agent who can be oneself or someone else. Shared representations are intersubjective. Rather than an egocentric space of actions, they constitute a "we-centric space" (Gallese 2003). Then we could say that action representations encode the parameter of the agent, but leave this parameter indeterminate and to be fulfilled. In brief, we do not need to start from the self in order to go to the other; rather we can start from what is common between oneself and the other.

*The Neutral Hypothesis: Mirror neurons are motor representations without any determinate agent:  $[x, \text{action}, \text{goal}]$ , with  $x$ , the variable of the agent to be fulfilled.*

The activation of mirror matching systems expresses this primary lack of differentiation, rather than the process of a covert simulation. We could similarly suggest that in emotional empathy, the emotion is by definition felt by a subject, but the subject is left indeterminate. The last step of the simulation process thus does not consist of inferring the target's emotion from the emotion felt by oneself, but in determining who the subject is (see Figure 3). Thus, the simulation process would imply a shared representation of emotion neutral in respect to the subject. Because of this primitive lack of differentiation, the emotional state can be attributed to oneself and to the other.

Descartes did not know that while he was looking at the people walking on the street, he activated the representation of walking from the first-person

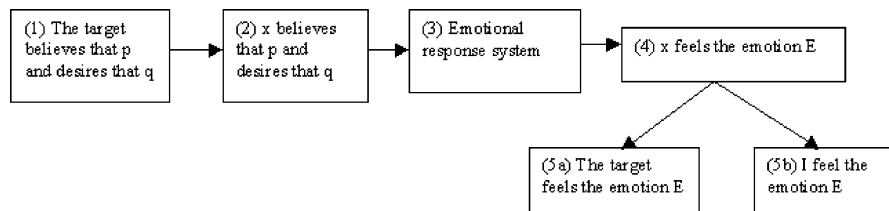


Figure 3. Neutral model of ST: a variable to fulfill.

perspective in his brain. The existence of sharing phenomena such as empathy or the activation of mirror neurons could not resolve his skepticism, as they do not provide any guarantee that others do exist. However, at the epistemic level, they argue against the radical difference between self-knowledge and knowledge of other minds. I have suggested two possible ways of interpreting them: either they reveal the simulation process underlying the knowledge of others (Gallese and Goldman 1998), or they stress the primary lack of differentiation in the representation of the self and the other.

### **Towards a new account of co-consciousness: A third interpretation**

According to the narrow interpretation of co-consciousness, we need to have access to what is in someone else's mind in the same way as we have access to our own mind. Two problems then arise. First, we do not have *de facto* such kind of knowledge: telepathy does not exist in our world. Second, the lack of epistemic difference between self-knowledge and the knowledge of other minds raises the question of how we can still recognize whose mental states we have access to.<sup>8</sup> According to a wider interpretation of co-consciousness, we know what the other feels because we feel the same. Co-consciousness then sounds more plausible and some phenomena such as empathy or, at the neurological level, mirror neurons seem to illustrate it. However, we may wonder whether it really addresses our original problem, whether it reduces the epistemic gap between the self and the other. We may indeed notice that co-consciousness in this sense is far from being infallible. I can mistakenly feel excited with my friend who is leaving for New York, while she actually feels anxious. This possibility of error is mainly due from the fact that my inference from my state to her state is invalid. It reveals the indirectness of the knowledge of other minds through co-consciousness. Even if I am able to report what someone else feels from a first-person perspective, I have still to derive what the other feels from what I feel. The knowledge of others is not as immediate as self-knowledge and the difference between them remains important. We seem then to reach a dead end.

It is possible that these interpretations of the phenomena of shared representations have gone wrong by giving priority to the self, rather than to intersubjectivity.<sup>9</sup> I have suggested a third interpretation of co-consciousness, which I have called the neutral account. Like co-consciousness in the narrow sense, you do not need to infer from you to the other; like co-consciousness in the wide sense, you do not need to read in someone else's mind. According to this new interpretation, one is co-conscious if one uses the same mental token for representing oneself and the other from a first-person perspective.

Gopnik suggests that we have the same representations of mental states for ourselves and for the others from a third-person point of view. However, this

does not sound plausible. Indeed, we may sometimes use the same abstract representations for ourselves and for the others, but we cannot eliminate the subjective perspective that accompanies self-knowledge, and more particularly in the case of emotions: there is something it is like to feel happiness from the point of view of the subject. What I am suggesting here is that rather than reducing self-knowledge to a specific kind of theoretical knowledge of other minds, as Gopnik tries to do, we should bring knowledge of others closer to self-knowledge: we share common representations of ourselves and the others from a first-person point of view. Mirror neurons do not represent actions from an abstract perspective, but rather from the intersubjective point of view. Thus, from a first-person perspective we are able to represent actions performed by someone else as if they were our own and there is no inference from you to the other involved here. This last interpretation may be more challenging for the privacy hypothesis denounced by Ayer. Obviously, we cannot reduce the knowledge of other people's actions to mirror neurons. However, we can suggest that these shared representations constitute a common epistemic basis of self-knowledge and the knowledge of other minds.

*The problem of the lack of differentiation*

However, one might argue that while the primary lack of differentiation between oneself and the other may underlie intersubjectivity, it also raises the problem of how we discriminate one's own states from the states of others. The neutral interpretation of co-consciousness would then meet the same problem as the narrow one. If there is indeed no difference between the representation of my states and your states, how can I differentiate them? There may even be a risk of confusing them and this conclusion challenges the usual assumption of "immunity to error through misidentification" (Shoemaker 1984). Intuitively, I could never believe that I am thinking while it is actually someone else who is thinking. I would directly know without any possible failure whether I am the thinker or not. Such immunity to error results from the complete asymmetry between the knowledge of one's own mind and the knowledge of other minds, which makes it impossible to confuse one's own thoughts with someone else's thoughts. But if we deny the existence of such asymmetry, as the neutral account does, then it becomes possible to be mistaken in state attributions. We may then wonder whether we should not give up the neutral interpretation.

It seems nevertheless that far from being problematical, the possibility of errors can be used as an argument for the neutral view. One can indeed be mistaken about the owner (or the agent) of one's own mental and bodily states, as it is particularly salient in schizophrenia. For instance, during verbal hallucinations, schizophrenic patients are talking to themselves but they think that someone else is talking; in delusions of control, they believe that they

control the actions performed by someone else or that their own actions are influenced by the will of other people. Similarly, some patients suffering frontal lesions automatically execute the action performed by someone else that they are observing, losing track of the distinction between their own intentions and the intentions of others (Lhermitte et al. 1986). Furthermore, it has been shown that even normal subjects can mistakenly self-attribute visually presented movements that they have actually not performed (Franck et al. 2001).

In summary, the primary lack of differentiation postulated by the neutral interpretation implies the necessity of a specific causal process that disambiguates the representations by articulating who the subject is. As Georgieff and Jeannerod (1998) say, the fact that action representations are shared is intimately related to the question of agency. Indeed, action representations by themselves do not specify who the agent is; we thus need a “Who” system, which determines whether the agent is the self or someone else (Georgieff and Jeannerod 1998; de Vignemont and Fournieret 2004). The possibility of misattribution confirms the existence of this process, which can breakdown. The primary lack of differentiation at the level of mirror neurons does not indeed imply that there is no difference at all between being the agent and being the observer. The “Who” system relies then on the comparison between: (i) visual and proprioceptive perception of the moving body and its effects on the environment; (ii) more putative central signals related to action initiation and planning (Frith et al. 2000). We may similarly suggest that we need a specific process dedicated to the distinction between one’s own emotions or thoughts and those of others.

The problem of other minds seems then to be completely reverted: the question is no longer how we can know the others, but rather how we can distinguish ourselves from the others. The questions of self-consciousness and intersubjectivity are indeed intimately related. In the perspective of the naturalization of self-consciousness, it would then be interesting to investigate the nature of the possible underlying cognitive mechanisms involved in the differentiation between the self and the other.<sup>10</sup>

It would seem then, that the problem might not be one of mapping the self onto the other so much as differentiating the self from the other (Gopnik and Meltzoff 1994, p. 191).

## Notes

1. I will not specify here the different versions of each of these conceptions.
2. His notion of co-consciousness has to be distinguished from William James’s notion. According to James, one person is said to be co-conscious of two different experiences if she is conscious of both of them as belonging to her own single stream of consciousness.



The question is then to understand the underlying basis of the unity of consciousness. Ayer is not concerned by this problem and appeals to the notion of co-consciousness when he addresses the problem of the privacy of the mental.

3. Briefly in his famous thought experiment, Jackson (1986) imagines that Mary has been brought up in a black and white environment without ever seeing any color. However, she is a scientist who knows every physical fact about colors. He then asks the following question: does Mary learn anything when she see for the first time a red tomato?
4. Ayer (1963) considers the possibility of feeling kinesthetic sensations from someone else's body but without describing the possible underlying neural details.
5. Unlike the narrow interpretation, there needs now to be two mental tokens that are of the same kind of states.
6. We are not concerned here by the knowledge of Peter's bodily position, but by the knowledge of his mental state.
7. Ayer (1963) mentions the multiple personality disorder as a possible case of co-consciousness, but rejects it as it does not clearly satisfy the condition (1).
8. However, we'll see that we may find a solution to this problem.
9. We may notice that Gallese (2003) is himself oscillating between the primacy of the self interpretation ("like me") and the neutral interpretation ("we-centric space").
10. I would like to thank Shaun Gallagher for his helpful comments.

## References

- Adams, F. 2001. Empathy, neural imaging and the theory versus simulation debate. *Mind and Language* 16(4): 368–392.
- Ayer, A. J. 1963. Privacy. In: *The Concept of a Person and Other Essays*, pp. 52–81. New York: St Martin's Press.
- Baron-Cohen, S. 1997. *Mindblindness*. Cambridge, MA: MIT Press.
- Calder, J. A., Keane, J., Manes, F., Antoun, N. and Young, A. W. 2000. Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience* 3(11): 1077–1078.
- Carr, L., Iacoboni, M., Dubeau, M. C., Maziotta, J. C. and Lenzi, G. L. 2003. Neural mechanisms of empathy in humans: A relay from neural systems for imitation in limbic areas. *Proceedings of the National Academy of Sciences of the United States of America* 100(9): 5497–502.
- Damasio, A. R. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace and Co.
- Davidson, D. 1987. Knowing one's own mind. *Proceedings and Addresses of the American Philosophical Association*, pp. 441–458.
- de Vignemont, F. and Fourneret, P. 2004. The sense of agency: A philosophical and empirical review of the 'Who' system. *Consciousness and Cognition* 13(1): 1–19.
- Decety, J., Grèzes, J., Costes, N., Perani, D. and Jeannerod, M. 1997. Brain activity during observation of action content and subject's strategy. *Brain* 120: 1763–1777.
- Descartes, R. 1641/1960. *Méditations Métaphysiques*. Paris: PUF.
- Eisenberg, N. and Strayer, J. 1987. *Empathy and Its Development*. Cambridge: Cambridge University Press.
- Frith, C. D., Blakemore, S. and Wolpert, D. 2000. Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society of London* 355: 1771–1788.
- Gallagher, S. 2001. The practice of the mind. Theory, simulation or primary interaction? *Journal of Consciousness Studies* 5–7: 83–108.

- Gallese, V. and Goldman, A. I. 1998. Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2: 493–501.
- Gallese, V. 2000. The acting subject: Towards the neural basis of social cognition. In: T. Metzinger (ed), *Neural Correlates of Consciousness: Empirical and Conceptual Questions*. Cambridge, MA: MIT Press.
- Gallese, V. 2001. The ‘Shared Manifold’ Hypothesis: From mirror neurons to empathy. *Journal of Consciousness Studies* 8(5–7): 33–50.
- Gallese, V. 2003. The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society of London, Series B* 358: 517–528.
- Georgieff, N. and Jeannerod, M. 1998. Beyond consciousness of external reality: A “Who” system for consciousness of action and self-consciousness. *Consciousness and Cognition* 7: 465–477.
- Goldman, A. I. 1992. Empathy, mind and morals. *Proceedings and Addresses of the APA* 66: 3.
- Goldman, A. I. 1993a. Consciousness, folk psychology and cognitive science. *Consciousness and Cognition* 2: 364–382.
- Goldman, A. I. 1993b. The psychology of folk psychology. *Behavioral and Brain Sciences* 16: 15–28.
- Goldman, A. I. 2002. Simulation theory and mental concepts. In: J. Dokic and J. Proust (eds), *Simulation and Knowledge of Action*, Advances in Consciousness Research, 45.
- Gopnik, A. and Astington, J. W. 1988. Children’s understanding of representational change and its relation to the understanding of the false belief and the appearance – reality distinction. *Child Development* 59(1): 27–37.
- Gopnik, A. 1993. How we know our minds: The illusion of first-person knowledge of intentionality. *The Behavioral and Brain Sciences* 16: 1–14.
- Gopnik, A. and Meltzoff 1994. Minds, bodies and persons: Young children’s understanding of the self and others as reflected in imitation and “theory of mind” research. In: S. Parker and R. Mitchell (eds), *Self-Awareness in Animals and Humans*. New York: Cambridge University Press.
- Gordon, R. M. 1996. Simulation without introspection or inference from me to you. In: P. Carruthers and P. Smith (eds), *Theories of Theories of Mind*. Cambridge: Cambridge University Press.
- Harris, P. 1989. *Children and Emotion*. Oxford: Blackwell Publishers.
- Hutchison, W. D. 1999. Pain related neurons in the human cingulate cortex. *Nature Neuroscience* 2: 403–405.
- Jackson, F. 1986. What Mary didn’t know. *The Journal of Philosophy* LXXXIII(5): 291–295.
- Jeannerod, M. 1994. The representing brain: Neural correlates of motor intention and imagery. *Behavioral Brain Sciences* 17: 187–245.
- Johnson, M. K., Hashtroudi, S. and Lindsay, D. S. 1993. Source monitoring. *Psychological Bulletin* 114: 3–28.
- Nisbett, R. E. and Wilson, T. D. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84: 231–259.
- Preston, S. D. and de Waal, F. B. M. 2002. Empathy: Its ultimate and proximate bases. *The Behavioral and Brain Sciences* 25(1): 1–20.
- Ramachandran, V. S. and Hirstein, W. 1997. Three laws of qualia : What neurology tells us about the biological functions of consciousness. *Journal of Consciousness Studies* 4: 429–457.
- Ravenscroft, I. 1998. What is it like to be someone else? Simulation and empathy. *Ratio* XI: 170–185.

- Rizzolatti, G., Fadiga, L., Gallese, V. and Fogassi, L. 1996. Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* 3: 131–141.
- Ryle, G. 1966. *Self-Knowledge*. Reprinted, 1994. In: Q. Cassam (ed), *Self-Knowledge*. Oxford: Oxford University Press.
- Shoemaker, S. 1984. Self-reference and self-awareness. In: *Identity, Cause and Mind*. Cambridge: Cambridge University Press.
- Wimmer, H. and Perner, J. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13(1): 103–128.
- Wispé, L. 1986. The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of Personality and Social Psychology* 50(2): 314–321.